HD28
.M414
no.416-69

WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

AUTOMATED JUDGMENTAL DECISION-MAKING

by

Norman E. Betaque* and G. Anthony Gorry**

416-69

September, 1969

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

AUTOMATED JUDGMENTAL DECISION-MAKING

by

Norman E. Betaque* and G. Anthony Gorry**

416-69

September, 1969

*Currently with U.S. Army, Viet Nam.

**Assistant Professor, Sloan School of Management, M.I.T.

HD28
.M414
no. 1120-80

Abstract

A utility theory model of the diagnosis and treatment of a serious medical condition, acute renal failure, has been implemented as part of a study of automated judgmental decision-making.

An interactive diagnostic computer program was used to test the model. This program employs formal decision procedures and the subjective assessment of likelihoods and preferences of experts to analyze the diagnostic/treatment problem.

The program's success in duplicating the decisions of expert clinicians in about 90 per cent of the cases used indicates that the method of analysis is not only a convenient structure for theoretically describing diagnosis and treatment, but that it is potentially a practical way of analyzing such decision problems.

RECEIVED
SEP 23 1969
M. I. T. LIBRARIES

1. Introduction

In this paper, some results of research into the automation of judgmental decision-making are reported. The specific problem we will discuss is a problem of medical diagnosis and treatment. This problem was investigated as part of a continuing research effort in computer-aided diagnosis [1, 2, 3]. Although the application we will discuss is a medical one, the approach we used has more general applicability, and perhaps its usefulness in problems of medical decision-making will stimulate interest in applying it elsewhere.

Experience with an interactive computer program for sequential diagnosis suggests that it is possible to implement programs which perform at high levels of competence in solving diagnostic problems [3]. In the areas in which it has been tested, the program performs at a level comparable to that of an expert human diagnostician. In this paper, an extension of the program to a new diagnostic-treatment problem is discussed. In the particular medical problem studied, that of acute renal failure, the determination of a therapeutic strategy is difficult for several reasons. First there are relatively few easily obtainable signs or symptoms which establish a definite diagnosis. Second, a number of the useful tests and treatments may involve considerable pain or risk for the patient. The decision problem is sufficiently difficult that the services of a diagnostic consultant are often required in dealing with a given patient. On the whole, the problem is a good one on which to test ideas about automating judgmental decision-making.

534168

In the next two sections, we will review the diagnostic program and the model upon which it is based and briefly sketch the nature of the acute renal failure problem. Then we will discuss the procedures we used to obtain from expert diagnosticians the data required by the program. Also we present some results of experiments using the program to select diagnostic and therapeutic strategies. Finally we offer some conclusions regarding the general usefulness of this approach.

2. A Computer Program for Sequential Diagnosis

In this study, we made extensive use of an interactive computer program for sequential diagnosis. Detailed discussions of this program are available in [2] and [4], and an example of its use is presented in [3]. Basically our efforts were directed at creating a data base for this program to use in assessing diagnosis-treatment strategies for acute renal failure.

Rather than discuss the details of the design and implementation of this program here, we will concentrate on the particular model of diagnosis upon which the program is based. Of particular interest will be the manner in which the subjective data provided by the expert clinicians is used. It should be noted that the diagnostic program remains unchanged from one application to the next. Only the data base used by the program reflects the given problem area.

The view taken in the design of the diagnostic program is that the selection of diagnosis-treatment strategies can be usefully viewed as a sequential decision problem under uncertainty.

The structuring of a decision problem for analysis initially involves the identification of available courses of action, their possible consequences, and the conditions under which those consequences might occur. Consider, for example, the diagnosis-treatment problem diagrammed by the decision tree in Figure 1. There are two available courses of action: treat for disease A or treat for disease B. There are three possible consequences: the patient either is cured, dies, or his condition remains unchanged. The conditions under which each of the consequences might occur is dependent not only on the course of action chosen, but on some unknown state of nature, in this case the particular disease the patient actually has. In this simple example, then, it is the unknown state of nature that introduces the element of risk into the problem. In real medical problems, of course, the consequences themselves may not be deterministic; the best treatment for a known disease may not guarantee a specific consequence. So, there is an element of risk even when the state of nature is known. A decision tree can be easily modified to include these situations by adding branches representing the additional consequences to appropriate nodes. Further extension of the problem can include provisions for improving the knowledge about the unknown state of nature by performing tests or experiments. In medicine, such tests might range from physical examination to major surgery. Whatever the case may be, the primary purpose of a test is to gain knowledge about the state of nature.

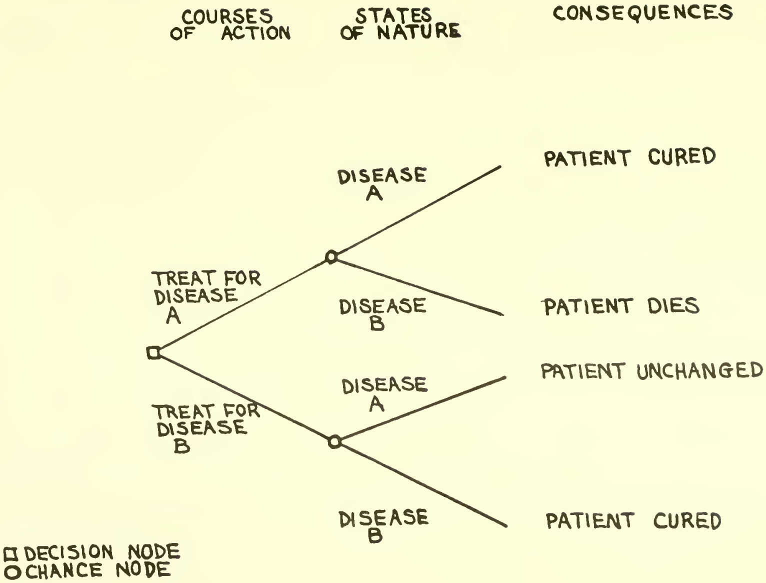
COURSES OF ACTION
STATES OF NATURE
CONSEQUENCES
TREAT FOR DISEASE A
TREAT FOR DISEASE B
DISEASE A
DISEASE B
DISEASE A
DISEASE B
PATIENT CURED
PATIENT DIES
PATIENT UNCHANGED
PATIENT CURED
□ DECISION NODE
○ CHANCE NODE

FIGURE I

Figure 2 illustrates the expanded decision tree.

Several aspects of the expanded problem are noteworthy. First, whether or not to perform a test is itself a decision having uncertain results. Obviously, if the results of a test are precisely predictable, it is senseless to perform the test. (If a test offers only therapeutic benefits it can be considered a treatment.) Note also that the ultimate consequences now include the costs of the tests. In some situations, the test costs are so small in relation to other consequences that they can safely be ignored; in others, however, the risk, patient discomfort, time, or inconvenience associated with a potentially useful test may lead a doctor to either carry out treatment without testing, or to give some other less informative, but less costly test. It should also be realized that the performance of a test will change the likelihood of a specific consequence occurring, as this is the whole purpose of testing: to reduce the risk of treatment by improving the knowledge about the unknown disease. It would seem desirable to conduct as many tests as possible to learn the true state of nature. Balanced against this desire, however, is the aggregate cost of testing, where cost may refer to time, pain, risk, and other factors as well as monetary expenses. At some point, the value of the knowledge attainable by further testing is exceeded by the costs. Hence it may be best to make the choice of treatments even though the actual disease is still, to some degree, uncertain.

It is in this process of weighing the merits of further testing,

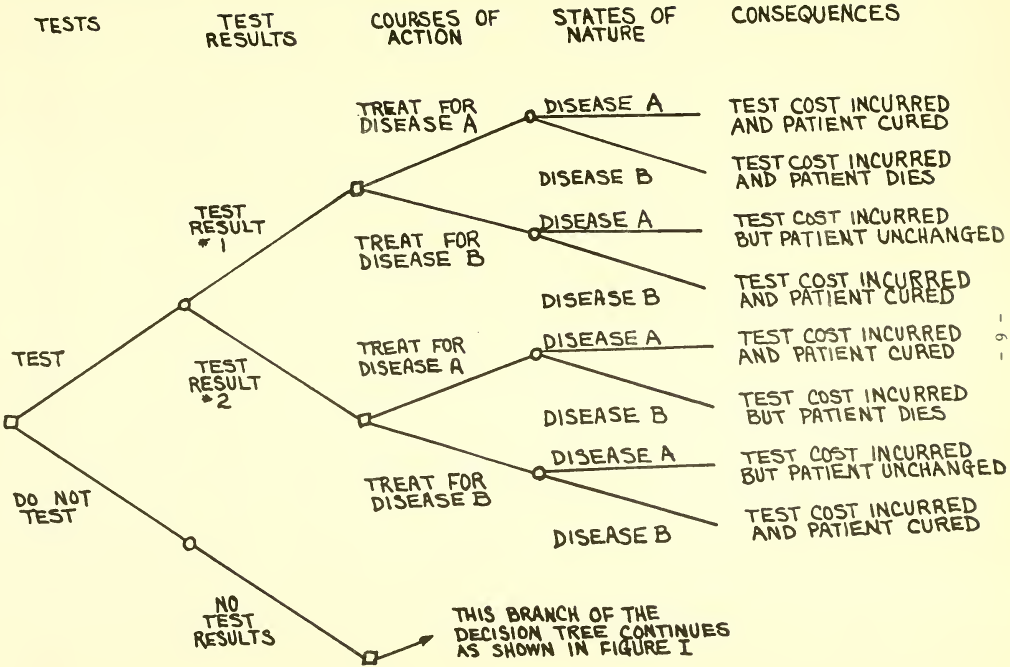
TESTS
TEST RESULTS
COURSES OF ACTION
STATES OF NATURE
CONSEQUENCES
TREAT FOR DISEASE A
DISEASE A
TEST COST INCURRED AND PATIENT CURED
DISEASE B
TEST COST INCURRED AND PATIENT DIES
TEST RESULT #1
TREAT FOR DISEASE B
DISEASE A
TEST COST INCURRED BUT PATIENT UNCHANGED
DISEASE B
TEST COST INCURRED AND PATIENT CURED
TEST
TREAT FOR DISEASE A
DISEASE A
TEST COST INCURRED AND PATIENT CURED
TEST RESULT #2
DISEASE B
TEST COST INCURRED BUT PATIENT DIES
TREAT FOR DISEASE B
DISEASE A
TEST COST INCURRED BUT PATIENT UNCHANGED
DISEASE B
TEST COST INCURRED AND PATIENT CURED
DO NOT TEST
NO TEST RESULTS
THIS BRANCH OF THE DECISION TREE CONTINUES AS SHOWN IN FIGURE I

FIGURE 2

re-evaluating the present state of knowledge, and finally choosing a course of action that a doctor invokes his own preferences and judgments. And it is these judgments and preferences which must be quantified to facilitate application of a computer to the decision problem.

These subjective assessments are an important input to the data base used by the program. Specifically, in addition to the identities of the various diseases to be considered, the program requires: 1) a set of decision alternatives (tests and treatments) with associated preferences; 2) a set of possible test results associated with various test-disease pairs; and 3) a set of possible consequences and preferences associated with treatment-disease pairs. In the model of diagnosis, both the associations (2) and (3) are probabilistic. The implied conditional probabilities along with the preferences for tests and treatments constitute the subjective data which are to be obtained from the clinician.

The basic operation of the diagnostic program given a data base for a particular problem is as follows. The user informs the program of the initial test and/or treatment results if any are available. Through the use of Bayes rule, the program computes the conditional probabilities for the diseases under consideration given these results.[1] In this calculation, the program employs: 1) _a priori_ disease probabilities (supplied by the

---

[1] If no initial results are available, the program begins with an _a priori_ probability distribution for diseases.

user), and 2) the conditional probabilities of the results given these diseases. The resulting distribution is called the program's *current view* of the problem, and the process by which the current view is obtained is called the *inference function*. Although the facility was not employed in this research, it should be noted that the inference function can deal with certain conditional dependencies among test and/or treatment results given disease [2].

The *test selection* function analyzes a decision tree like that in Figure 2 to select either a test or a treatment for the user to employ. For most real problems, the decision tree is very large, and the development of optimal diagnostic strategies is computationally infeasible. In addition, changes and reassessments of the subjective data employed by the program make inadvisable the investment of a large computational effort to analyze the tree for a given set of values. Hence the program uses heuristics to restrict the growth of the decision tree during analysis. These heuristics can vary from problem to problem. Some of the commonly used ones are discussed in [2], and those used in the renal failure problem will be described in Section 3.

In any event, once the user has performed the test or treatment selected, the results are processed by the inference function to obtain a new current view. Then the test selection function again is invoked.

The program has other capabilities [2], but those of the inference and test selection functions are of principal interest here. It is used in an interactive manner, and the basic purpose is to develop good

diagnostic-treatment strategies through the continued re-evaluation of decision alternatives in the light of recently accumulated evidence.

3. Modelling the Renal Failure Problem

3.1 The Basic Model

The syndrome, acute renal failure, was selected for several reasons. First, the diagnosis of the cause of the acute renal failure can be a difficult problem. It is often necessary to try a treatment before a definite diagnosis has been confidently determined. Second, there are a number of tests available to a doctor which can be useful in the diagnosis, but with which are also associated varying degrees of risk, discomfort, and other sacrifices categorizable as test costs. Third, the diseases possibly responsible for the syndrome are very diverse in their seriousness and their responses to available treatments. In some cases, neither the best available treatment nor any of a number of mistreatments is likely to significantly affect the course of the disease. On the other hand, there are situations in which correct diagnosis and treatment is likely to be very successful, but mistreatment can result in grave consequences. Last, but most important, acute renal failure was chosen because there were doctors available who were not only experienced in the diagnosis and treatment of the syndrome, but had expressed an interest in the project and a willingness to contribute their time and expertise.

An effort was made to keep the model relatively uncomplicated. Only tests thought to have non-trivial costs or to be useful in very specific situations were included. Routine tests and questions pertaining to

medical history or physical examination were not considered. Also similar treatments or consequences were consolidated. For example, although a drug can be administered in various doses, only one dose of each drug was considered. Similarly all treatment consequences were placed in one of three categories.

The model developed for the analysis of acute renal failure is diagrammed in Figure 3. The diseases considered, the tests and their possible results, the treatments, and the consequences are presented in Tables 1 through 4. Note that Figure 3 really represents an entire decision tree. A test need not be given before a treatment can be chosen (Part of the analysis is to determine whether or not to give a test at all.). Similarly, the model does not necessarily restrict the analysis to one stage of testing as shown in the figure. A sequence of any desired number of testing stages can be evaluated before choosing a course of action. So, the model is basically the same as that described earlier in Figure 2.

Several simplifying modifications, however, have been made. First, the test and treatment costs were considered to be independent of the actual disease as well as each other. Therefore, these costs are effectively predetermined tolls which must be paid at the time a test is given or treatment is carried out, rather than additional consequences of a testing-treatment policy. This modification greatly simplified data collection and actual computation since the decision tree beyond the treatment decision node remains the same regardless of testing accomplished. Likewise, the cost of any particular test or treatment was constant regardless of previous or subsequent testing.

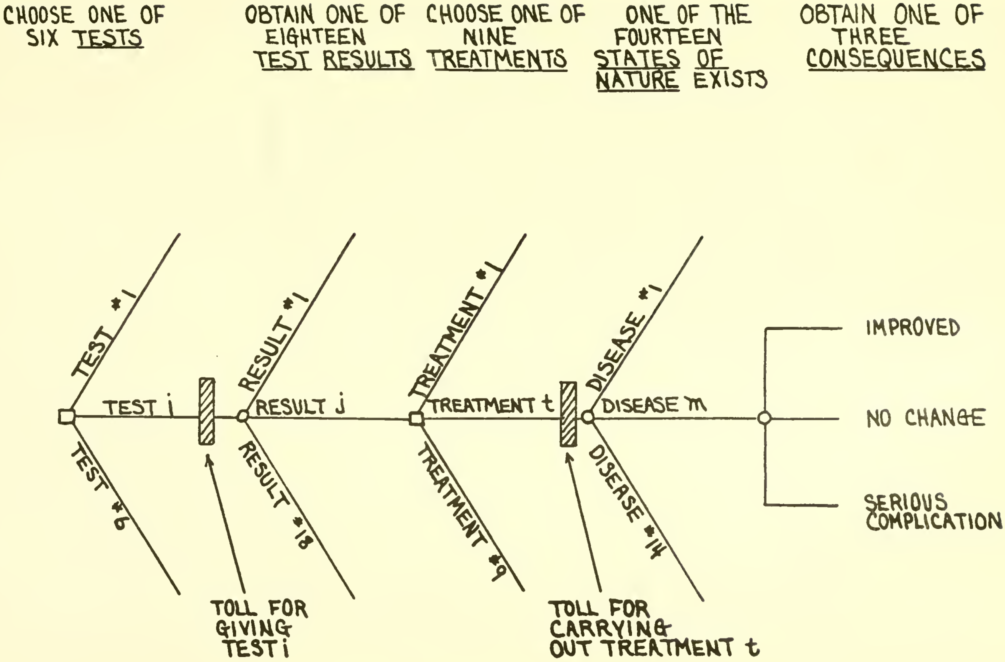
CHOOSE ONE OF SIX TESTS
OBTAIN ONE OF EIGHTEEN TEST RESULTS
CHOOSE ONE OF NINE TREATMENTS
ONE OF THE FOURTEEN STATES OF NATURE EXISTS
OBTAIN ONE OF THREE CONSEQUENCES
TEST #1
TEST i
TEST #6
RESULT #1
RESULT j
RESULT #18
TREATMENT #1
TREATMENT t
TREATMENT #9
DISEASE #1
DISEASE m
DISEASE #14
IMPROVED
NO CHANGE
SERIOUS COMPLICATION
TOLL FOR GIVING TEST i
TOLL FOR CARRYING OUT TREATMENT t

FIGURE 3

TABLE 1

DISEASES

	DISEASES	ABBREVIATION
D01	Acute Tubular Necrosis	ATN
D02	Functional Acute Renal Failure	FARF
D03	Obstruction of Urinary Tract	OBSTR
D04	Vasculitis	AG
D05	Cortical Necrosis	CN
D06	Hepatorenal Syndrome	HS
D07	Pyelonephritis	PYE
D08	Atheroembolism	AE
D09	Renal Infarction	RI
D10	Renal Vein Thrombosis	RVT
D11	Acute Interstinal Nephritis	AIN
D12	Scleroderma	SCL
D13	Acute Renal Disease/Chronic Disease	ARDCD
D14	Malignant Hypertension	MH

TABLE 2

TREATMENTS AND CONSEQUENCES

TREATMENTS

T01	No Therapy
T02	IV Fluids
T03	Mannitol
T04	Surgery for Obstruction
T05	Steroids
T06	Antibiotics
T07	Surgery for Clot
T08	Antihypertensive Drugs
T09	Heparin

CONSEQUENCES

R01	Patient's Condition Improved
R02	No Change in Patient's Condition
R03	Patient Experiences a Serious Complication

TABLE 3

TESTS

	TESTS	POSSIBLE RESULTS*
T10	Biopsy	R04 thru R15
T11	Retrograde	R16, R17
T12	Plain Film of Abdomen	R18, R19
T13	Catheterization	R16, R17
T14	Arteriography	R20, R21

*See Table 4 for listing of test results.

TABLE 4

TEST RESULTS

TEST RESULTS

R04	No Specific Finding
R05	ATN
R06	AG
R07	CN
R08	HS
R09	PYE
R10	AE
R11	RI
R12	AIN
R13	SCL
R14	ARDCD
R15	MH
R16	OBSTR
R17	NO OBSTR
R18	Enlarged Bladder Visible
R19	Bladder Not Visible
R20	Blood Clot
R21	No Blood Clot

In addition, it was assumed that in any specific case, the syndrome is actually a result of only one disease; the model does not allow the patient to be suffering from more than one disease at a time. Similarly, there is no provision for simultaneously carrying out more than one treatment. Both of these assumptions simply limit the size of the model without affecting its basic structure.

Further simplification was achieved by assuming that the preferences of the three possible consequences of any treatment not only reflected the patient's condition subsequent to the treatment, but summarized all future costs and consequences. Hence, in selecting a course of action, the model used the consequence of a treatment as the horizon for the analysis. In real life, of course, the problem might not terminate at that point. If a treatment results in other than improvement in the patient's condition, there is still a diagnosis-treatment decision confronting the doctor. To consider such a situation using the present model, the a priori probabilities of the diseases would have to be updated to reflect the experience of the ineffective treatment and the problem recycled, once again determining a testing-treatment policy.

One aspect of the costs associated with tests and treatments requires amplification. Costs can be conveniently divided into risk and non-risk categories. The category of risk is the probability that a serious consequence will develop pursuant to the administration of a specific test or treatment. The category of non-risk may include monetary costs,

patient discomfort, time, hospital facilities used, or whatever other non-risk factors the doctors may wish to consider. It should be realized that the risk of testing is a consequence, not a test result; it does not provide information about the cause of the acute renal failure since the test costs are being considered independent of the actual disease. Thus, the toll for performing a test must include both risk and non-risk associated categories of cost. The risk associated with a treatment, however, cannot be considered independent of the actual disease, but is, in fact, reflected in the conditional probabilities obtained from the doctor: his judgment of the likelihood of each consequence occurring given each combination of diseases and treatments. Thus, the tolls associated with treatments need only consider the non-risk category of costs.

In order to use the diagnostic program to solve this model, we needed 1) probabilities and 2) preferences for tests, treatments, and consequences expressed on some common scale.

3.2 The Probabilities

All the probabilities used in this study were strictly subjective estimates made by one clinician. There was no study of historical data. Except for obvious instances of mutual exclusion, no attempt was made to assess conditional dependencies.

The doctor was simply asked to directly estimate the required probabilities based on his own best judgment. If he was aware of statistical studies that had substantiated some of the data, he was

cautioned to use the results of those studies only if he really believed them. The purpose of subjective probabilities was to quantify the doctor's own judgments, whether these judgments were accurate or not.

There are three types of probabilities required for implementation of the model: the *a priori* (initial) probabilities of the diseases, the conditional probabilities of the test results given the tests and diseases, and the conditional probabilities of the consequences given the treatments and diseases. The *a priori* probabilities are the inputs to the model and change according to the specific medical case that is being processed. They represent the sum total of all information gained about a patient prior to the analysis of the model; the data from the patient's medical history, results of routine tests, findings from a physical examination, and the likelihood of the possible diseases occurring are all reflected in the *a priori* probabilities.

The conditional probabilities are a semi-fixed part of the model. They can be changed if experience warrants a revision, but they normally form a data base which remains consistent from case to case. The conditional probabilities of the test results given the tests and diseases represent the likelihood that any patient who has a specific disease and is administered a specific test will demonstrate each of the possible test results. This information is indicative of the reliability of a test and must reflect the possibility of human error as well as test inadequacies. The conditional probabilities of consequences given the treatments and diseases represent the likelihood that any patient who has a specific

disease and is administered a specific treatment will experience each of the possible consequences. It is a measure of the effectiveness of each of the treatments.

Eliciting of conditional probabilities from the doctor took only minutes. There were, however, several interesting aspects of this portion of the project. The doctor who provided all the subjective probabilities had previous experience with both computers and their application to problems in medical diagnosis. He also had worked with subjective probabilities and had some intuitive understanding of Bayes' Theorem. He was aware that a zero probability expressed a judgment of impossibility and, therefore, eliminated certain options with no opportunity for recall later in the case. Thus, a zero probability could exclude from consideration one of the possible diseases. The result of this understanding was the estimation of many probabilities at .01. This meant that those test results or consequences being considered were effectively not possible except through accident or freak circumstances, but he wished to keep open his options for later in the problem.

The doctor's approach to assigning probabilities to his judgments was usually to identify the possible but unlikely occurrences, assign low probabilities to them, then allot the remaining probability to the most likely occurrence. Sometimes the unlikely occurrences were first considered as a group, a probability assigned to that group, then later divided among the individual events. This accounts for such subjective probabilities as .87, .29, .09, etc.

It was also discovered that in some situations the doctor felt he could not make the required judgment with any great degree of confidence. These situations arose primarily when considering the likelihood of consequences occurring as a result of mistreatment. The doctor's reaction was that if a patient had the disease specified, the particular treatment being considered was not appropriate and had never, to his knowledge, been attempted. Obviously, any doctor who did grossly mistreat a disease would not be anxious to acknowledge that he had done so, let alone document and publish the case. So, there were times the doctor was simply asked to guess. The rationalization behind this approach was that doctors, either consciously or unconsciously, make similar judgments daily, and their guess should be at least as good as anyone elses.

3.3 Preferences and Utility Theory

To establish the required preferences for the tests, treatments, and consequences in the model, we employed a lottery technique based on utility theory. Utility theory is an axiomatic theory of decision making under risk. It is a prescriptive theory; it specifies the course of action a decision maker should choose to be consistent with his preferences and judgments. In short, it states that if a decision maker accepts the axioms of the theory, he should choose that course of action which maximizes his expected utility, utility being a numerical scale over the decision maker's preferences.

The axioms of utility theory are fundamental assumptions which form the basis of the theory and from which can be derived all other aspects of the theory. On the surface, it appears that any man trying to act rationally would not only accept the axioms but would like to make all his decisions consistent with them. In reality, they are not as universally attractive as would first appear. In fact, there are several sets of axioms leading to similar variations of utility theory. Luce and Raiffa [5] discuss some of the psychological implications of accepting a utility theory as well as the detailed mathematical statement of the axioms. Presented here is one set of axioms, or assumptions, taken from Chernoff and Moses [6].

Assumption 1. (Ordering of Alternatives) An individual faced with two prospects, A and B, will be able to decide whether he prefers A to B, B to A, or is indifferent between them.

Assumption 2. (Transitivity) If prospect A is preferred to B, and B is preferred to C, then A is preferred to C.

Assumption 3. If prospect A is preferred to prospect B, which is in turn preferred to C, then there is a mixture*of A and C which is preferred to B, and a mixture of A and C to which B is preferred.

Assumption 4. If prospect A is preferred to prospect B, and C is another prospect, then any mixture of A and C will be preferred to the same mixture of B and C.

---

*The nature of this mixture will be discussed later.

The four assumptions are sufficient to establish a linear scale, called a utility scale, over the prospects. That is, to each prospect corresponds a number called the utility of that prospect. This utility scale has the following properties:

1. The utility of prospect A is greater than that of prospect B if and only if A is preferred to B.

2. If B is a prospect where, with probability p, the individual faces prospect A and with probability 1-p he faces prospect C, then

$$\text{Utility of B} = p(\text{Utility of A}) + (1-p)(\text{Utility of C})$$

Property 1 merely says that more preferred prospects have higher utilities. Property 2 states that if prospect B can be considered a lottery offering a p chance at prospect A and a 1-p chance at prospect C, then the utility of B is the expected utility of the lottery.

The computations made in applying utility theory to a decision problem stem directly from these two properties. First, it is necessary to assess utilities over all the consequences being considered. Next, if the likelihood of prospects occurring can be expressed as probabilities, it is possible, by working backwards through the decision tree and invoking Property 2, to determine the expected utility of each course of action at each decision node. According to Property 1, then, the decision maker should choose, at each decision node, the course of action promising the highest expected utility.

There are three categories of preferences over which utilities are to be assessed: the preferences for test costs, which include both the risk and non-risk aspects; the preferences for treatment costs, which include only the non-risk aspects; and the preferences for the consequences. Ultimately, these preferences must be quantitatively expressed on a common utility scale. Initially, however, the approach was to establish utility scales over each of the categories individually, then to seek a way of merging them into a common scale.

The technique for assessing utilities is derived directly from the third assumption of utility theory. The mixtures referred to in Assumption 3 are probabilistic in nature. That is, they are similar to lotteries; they represent a chance at attaining prospect A and the complementary chance at attaining prospect C. If the lottery is greatly biased towards prospect A (meaning that there is a sufficiently good chance of attaining prospect A) an individual would prefer the uncertainty of the lottery to the certainty of prospect B, since the order of preference is A - B - C. On the other hand, if the lottery is sufficiently biased towards the least preferred prospect C, an individual would prefer the certainty of prospect B. Presumably, then, there is some lottery of prospects A and C such that an individual is indifferent in his preference between the certain prospect B and the lottery.

If the upper and lower limits of a utility scale are arbitrarily set and assigned to prospects A and C respectively, the utility of prospect B can be computed using the provisions of Property 2 of a utility scale.

Thus, a scale can be constructed which reflects the relative preferences of prospects A, B and C. Similarly, any prospect for which the preference falls between the extremes of prospects A and C can be represented on the scale by identifying the appropriate indifference lottery of prospects A and C.

The application of this technique of utility assessment to preferences relevant to the medical diagnosis-treatment problem primarily involves posing the appropriate lottery choices in such a manner that only the desired preferences are involved. For example, in trying to assess utilities over preferences for treatments, it has been explained that only the non-risk aspects of treatment tests should be considered. A doctor, however, in weighing the merits of a proposed treatment, does not separate the treatment costs into risk and non-risk categories. Nor does he divorce preferences for the costs as a whole from other aspects of the case he is considering. He invokes his training, experience, and intuition to simultaneously evaluate all factors bearing on the specific case at hand. Thus, to assess utilities, it is necessary to pose questions which enable the doctor to mentally strip away all non-pertinent considerations and focus only on issues material to the preferences desired.

3.4 Utility Assessments

In this section, we will review the procedure by which utility assessments were made. A number of interesting details will be excluded from our discussion because of space limitations. The interested reader is referred to [7] for a complete discussion of procedures and problems.

The procedure used to establish the utility of tests and treatments was as follows. First the non-risk costs were considered. The most preferred and least preferred treatment (test) were identified. For treatments, the doctor was asked

> "If all treatments were guaranteed to cure the patient, regardless of the actual disease, would you prefer the specific treatment in question, or a p chance of "surgery (clot)" (the least preferred treatment and a 1-p chance of "no therapy" (the most preferred treatment)?"

A similar question was employed for tests.

Notice that the question was posed so that it would be clear to the doctor that neither the effectiveness nor the risk of the treatment (test) was to be considered. To assess the risk costs of tests, the doctor was asked to estimate the probability that each test, independent of the disease, would lead to a "serious complication". The utility of each test was taken as the assessed utility of its non-risk cost plus this probability times the utility of "serious complication".

It was relatively easy to establish a utility scale for the three consequences using a similar lottery technique. Then except for the addition of the risk costs to the test utilities, the three scales were complete. Certain problems, however, were encountered in attempting to merge these scales into one.

The doctors involved in this study had little difficulty in comparing tests and treatments, in spite of the fact that test costs include risk and treatment costs do not. Apparently because both tests and treatments involve doing something to the patient, the factors considered in evaluating costs are similar. Merging consequences into the common scale was much more difficult, because to the doctors there seemed to be no ground for comparison. Certainly the consequence "improved" is preferable to all tests and treatments, and all tests and treatments are preferred to "serious complication", but there was no obvious way of relating the preferences for consequences to the common scale for tests and treatments.

After an unsuccessful attempt [7], the following approach was devised. The doctor was asked to make a choice between two courses of action. One action offered an opportunity to carry out "surgery (clot)" with the knowledge that the treatment would result in the consequence "improved". In other words, there was no risk associated with the surgery; only the non-risk costs would be incurred. The alternative to surgery was to accept a p chance at "serious complication" and the complementary chance, 1-p, at "improved". The p was then adjusted until the doctor expressed indifference between the two courses of action. The result, p equal to .10, was interpreted to mean that the utility of "surgery (clot)" should be about one tenth the utility of "serious complication". Although the same procedure was applied to other treatments in an attempt to establish several reference points for merging the two utility scales, all values of p were so close to zero that no useful differentiation was obtained. Accordingly, the lower extremes of both scales were considered to have negligible utility.

4. Experimentation with the Model

In order to test the model discussed above, a series of hypothetical decision problems were formulated for solution both by doctors and by the diagnostic program. We were fortunate in that two expert clinicians with extensive experience with the renal failure problem agreed to participate in this experiment.

The experimental procedure was as follows. The required conditional probabilities were supplied by doctor #1. These probabilities were used in experiments with doctor #2 as well.[1] A utility scale was established for each physician. Then each doctor was asked for his decision in each of a set of 28 hypothetical decision problems. A decision problem is the selection of a test or treatment given an *a priori* probability distribution over the diseases in question. The program solved the same decision problems using the two utility scales respectively. The decisions reached by the doctors and by the program using their utility scales are presented in Table 5.

From this table, it is apparent that the program using the decision model and the subjective inputs from the clinician does well in replicating their decisions. In 51 of the 56 decision problems, the program's decision was the same as the clinician's or was judged by him to be equally good. Notice that the program agreed as closely with the clinician whose utility

---

[1] In two instances, it was discovered that doctor #2 disagreed with one of these probabilities, and it was changed to reflect his belief.

Table 5

Experimental Results

Case Number	Distribution	Dr. #1	Model Utility Scale (#1)	Dr. #2	Model Utility Scale (#2)
C01	OBSTR .75 ATN .12 AG .13	T11	T11	T11	T11
C02	FARF .90 ATN .05 AG .05	T02	T02	T02	T02
C03	ATN .85 FARF .10 AG .05	T01	T01	T01	T01
C04	ATN .50 OBSTR .50	T11	T11	T11	T11
C05	MH .75 ARDCD .20 SCL .05	T08	T08	T08	T08
C06	AG .75 MH .25	T05	T05	T08	T08
C07	MH .50 AG .50	T10**	T08	T08	T08
C08	RI .50 CN .50	T14***	T09	T09	T09
C09	PYE .80 AG .20	T06	T06	T06	T06
C10	PYE .40 AG .40 AE .20	T10	T10	T10	T10
C11	FARF .50 ATN .50	T02*	T10	T02	T02
C12	AG .77 ARDCD .15 MH .08	T05	T05	T05	T05
C13	PYE .60 ATN .30 OBSTR .10	T06	T06	T11	T11
C14	PYE .10 OBSTR .60 ATN .30	T11	T11	T11	T11
C15	HS .60 ATN .20 OBSTR .20	T11	T11	T11	T11
C16	RI .50 CN .20 AG .30	T14	T14	T09	T09
C17	SCL .50 MH .50	T08	T08	T08	T08
C18	RVT .50 ATN .30 FARF .20	T09***	T02	T09	T09
C19	ATN .50 AE .50	T01	T01	T01	T01
C20	ATN .35 FARF .35 AE .30	T02*	T10	T10	T10
C21	OBSTR .05 AG .50 AE .45	T12	T12	T11*	T12
C22	SCL .10 ATN .25 OBSTR .65	T11	T11	T11	T11
C23	SCL .10 CN .30 AG .60	T05	T05	T05	T05
C24	SCL .50 CN .50	T09	T09	T09	T09
C25	RVT .20 SCL .20 ATN .60	T01	T01	T01	T01
C26	MH .50 PYE .40 RI .10	T06 & T08*	T10	T06 & T08*	T10
C27	FARF .30 ATN .50 OBSTR .20	T11	T11	T11	T11
C28	FARF .50 ATN .10 OBSTR .40	T11	T11	T11	T11

* The model and the doctor disagree.
** The doctor later changed his decision to agree with the model.
*** The doctor stated that either his or the model's decision was appropriate.

scale it employed as did the other clinician. It is interesting to note that in the cases in which the doctors disagreed on the appropriate course of action, the model followed the appropriate doctor (that is the one who had supplied the utilities).

Several factors are important in the interpretation of these results. First, the two doctors have worked closely together for a number of years, and they share similar views about the renal failure problem. Also there are many factors which enter into a doctor's decision -- such as patient identification, patient condition, and duration of treatment -- which were not included in the model. For this reason, the model was tested on hypothetical rather than actual cases. Therefore, caution is appropriate in assessing the significance of these results.

On the other hand, none of the model's limitations seemed to be inherent in the characterization of medical diagnosis as decision making under risk. On the contrary, the doctors felt that this is a good way of approaching medical decisions. After all, the procedure of consciously identifying various courses of action, their known costs, their risks, and their possible consequences is simply a very methodical way of accomplishing the same type of analysis that doctors invoke intuitively. The primary difference is that doctors do not try to quantify either their judgments or preferences. Yet, the doctors who participated in this project were able, with only several exceptions, to readily provide the information necessary to make those quantifications.[2]

---

[2] A more complete discussion of the medical significance of this effort is available in [7].

5. Conclusions

The results of our study of the use of a computer to make judgmental decisions in the diagnosis and treatment of acute renal failure problems are encouraging for several reasons. On the one hand, a potential for dealing with the specific medical problem has been demonstrated. In addition, some insights have been gained into the usefulness and difficulties of the approach in a wider class of decision problems.

The management of the acute renal failure syndrome is an important medical problem. Although the incidence of the problem is relatively small in the context of all disease treatment problems, the potential risk to the patient's life is sufficiently great to give the problem special significance. In addition, the determination of good diagnostic and treatment strategies is a complex decision problem. The best solution, of course, is to employ the services of an expert clinician with extensive experience in this area. Unfortunately, the number of such experts is very limited.

The performance of the average doctor on this problem is apt to be significantly poorer than that of such an expert. Hence it is important to supplement and amplify the capability of these doctors to deal with the renal failure problem. The diagnostic program with the appropriate data base (formed from the judgments of experts) represents a distributable expertise. Although it is not a substitute for the expert, it is potentially superior to the average doctor in its decision making capability.

The use of directly estimated subjective probabilities is an area in which further research might be helpful. There was no effort in this project to confront the many issues relevant to probability estimation or the use of subjective probabilities. The doctor merely stated his estimates, and they were thereafter faithfully used in the model. It was pointed out, though, that one doctor had previous experience in estimating probabilities. So, there was some prior indication that this approach would prove adequate, and the results seem to substantiate that assumption.

A close look at the overall effect of the probabilities suggests that this casual approach of direct estimation may be a good technique. Even for this limited project it would have taken a major effort, over a long period of time, to statistically determine the required probabilities with any high degree of confidence. On the other hand, direct estimation took only a few minutes. And even though the doctor's estimates may not be individually accurate, he usually knows whether a probability should be high, low, or somewhere in the middle. As long as the numbers he chooses reflect his judgment about the relative effectiveness of the tests or treatments on each of the diseases, the fact that he has difficulty distinguishing between a 0.20 and a 0.25 does not significantly affect the model. As a practical matter, should such small variations in the probabilities result in a change in the recommended test or treatment, a doctor would probably agree that either decision is appropriate for the case.

The situations in which direct estimation seemed least satisfying were those concerning rare, unusual or unheard of mistreatments. The doctor simply did not know what would happen if a patient were so grossly mistreated; so his probability estimates were just guesses. Even these situations had their contribution to the project: they helped identify areas in which medical knowledge needs to be improved. Experiments or statistical studies in these few areas may be all that is necessary to raise a doctor's confidence in some of his guesses and provide entirely adequate probabilities for utility theory type models.

The direct use of utility theory to establish scales over tests, treatments and consequences also had its successes and failures. In scaling the non-risk costs of tests and treatments, the lottery technique was very satisfactory. The doctors had no trouble responding to the questions and were satisfied with the resulting scales. In fact, the practice of grouping several tests and treatments having similar non-risk costs proved quite appropriate to the type analysis a doctor makes. The key issue in these situations was whether, within a group, the non-risk cost would affect a doctor's decision. Even though there may be differences in non-risk costs among the tests or treatments in a group, and even though those differences may be very apparent when isolated in a lottery choice situation, as a practical matter, the risk or effectiveness of the tests or treatments so greatly overshadows their differences in non-risk costs that those non-risk costs never really affect a decision. In fact, the risks associated with the various tests considered in this project sufficiently overshadowed

the non-risk costs that those costs might safely be ignored. More experience in dealing with medical decisions might lead to other methods for simplifying utility assessment.

Attempts to apply the lottery technique to the three consequences and to then relate those consequences to the tests and treatments proved frustrating. This project never did establish a satisfactory means of quantifying those relationships. The stumbling block was the extreme disparity between the undesirability of the "serious complications" and that of the non-risk costs of the treatments. Isolating the "serious complications" from a real medical situation and focusing on their undesirability seemed to magnify their gravity. When efforts were made to quantify this severe attitude, the effect was the establishment of a very cautious model, one which would recommend treatment only after a very sure diagnosis had been obtained.

The trade-off operating here is that the higher the utility of a "serious complication", the more the probabilities, hence the risk, dominate the model. As the utility of "serious complication" is lowered, the more the relative non-risk costs influence the decision. In this project, both the computer program used and the on-line mode of operating the computer made it easy to observe this trade-off. Utilities could be easily changed and a case re-run on the model. Immediate response from the computer allowed convenient comparison of present and past performance

of the model and evaluation of the affect of the changed utilities. It also increased intuitive understanding of the model and enabled the users to identify means of improving the model's performance.

Because there were only three consequences involved, this man-machine interaction was sufficient for adjusting consequence utilities until the model seemed to work. In more complex models, however, it may not be possible to adequately represent all consequences in so few categories, and trial and error methods might prove inadequate. A useful approach might be to define a category of consequences that is comparable to treatments, and use it both to establish a meaningful scale over the consequences and to quantify the relationship between consequences and treatments. In any event, this comparison is considered to be the least satisfying aspect of this project. Any improvement in the techniques used to make the comparison would greatly enhance the appeal of using utility theory to model medical diagnosis-treatment problems.

REFERENCES

[1] Gorry, G. Anthony, "A System for Computer-Aided Diagnosis", Project MAC Technical Report, MAC-44, 1967.

[2] Gorry, G. Anthony, "Strategies for Computer-Aided Diagnosis", Mathematical Biosciences, 2, (1968), pp. 293-318.

[3] Gorry, G. Anthony, and G. Octo Barnett, "Sequential Diagnosis by Computer", Journal of the American Medical Association, Vol. 205, September 16, 1968, pp. 849-854.

[4] Gorry, G. Anthony, and G. Octo Barnett, "Experience with a Model of Sequential Diagnosis", Computers in Biomedical Research, Vol. 1, No. 5, pp. 490-507.

[5] Luce, Duncan R., and Howard Raiffa, Games and Decisions: Introduction and Critical Survey, New York: John Wiley and Sons, Inc., 1968.

[6] Chernoff, Hermand, and Lincoln E. Moses, Elementary Decision Theory, New York: John Wiley and Sons, Inc., 1959.

[7] Betaque, Norman E., "Utility Theory Applied to Medical Diagnosis and Treatment", Electrical Engineering Master's Thesis, M.I.T., May, 1969. Also Working Paper No. 393-69, Sloan School of Management, M.I.T., 1969.

26 '72

R 22 '74

Date Due

MAR 17 '77
MAY 11 '78
JUN 8 '78
DEC 4 1980
OCT 3 '81

Lib-26-67

MIT LIBRARIES
3 9080 003 874 861
410-69

MIT LIBRARIES
3 9080 003 874 853
411-69

MIT LIBRARIES
3 9080 003 874 846
412-69

MIT LIBRARIES
3 9080 003 874 838
413-69

MIT LIBRARIES
3 9080 003 874 812
414-69

MIT LIBRARIES
3 9080 003 874 820
415-69

MIT LIBRARIES
3 9080 003 874 804
416-69

MIT LIBRARIES
3 9080 003 869 879
417-69

MIT LIBRARIES
3 9080 003 874 622
418-69

MIT LIBRARIES
3 9080 003 901 664
418-69A